atmosphere

MDPI

Review

Data-Driven Weather Forecasting and Climate Modeling from the Perspective of Development

Yuting Wu [1,2] and Wei Xue [1,2,3,*]

[1] School of Computer Science and Technology, Qinghai University, Xining 810016, China; yutingwu@qhu.edu.cn
[2] Qinghai Provincial Laboratory for Intelligent Computing and Application, Qinghai University, Xining 810016, China
[3] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
* Correspondence: xuewei@mail.tsinghua.edu.cn

Abstract: Accurate and rapid weather forecasting and climate modeling are universal goals in human development. While Numerical Weather Prediction (NWP) remains the gold standard, it faces challenges like inherent atmospheric uncertainties and computational costs, especially in the post-Moore era. With the advent of deep learning, the field has been revolutionized through data-driven models. This paper reviews the key models and significant developments in data-driven weather forecasting and climate modeling. It provides an overview of these models, covering aspects such as dataset selection, model design, training process, computational acceleration, and prediction effectiveness. Data-driven models trained on reanalysis data can provide effective forecasts with an accuracy (ACC) greater than 0.6 for up to 15 days at a spatial resolution of 0.25°. These models outperform or match the most advanced NWP methods for 90% of variables, reducing forecast generation time from hours to seconds. Data-driven climate models can reliably simulate climate patterns for decades to 100 years, offering a magnitude of computational savings and competitive performance. Despite their advantages, data-driven methods have limitations, including poor interpretability, challenges in evaluating model uncertainty, and conservative predictions in extreme cases. Future research should focus on larger models, integrating more physical constraints, and enhancing evaluation methods.

Keywords: data-driven model; deep learning; weather forecasting; climate modeling; Numerical Weather Prediction (NWP)

check for updates

Citation: Wu, Y.; Xue, W. Data-Driven Weather Forecasting and Climate Modeling from the Perspective of Development. *Atmosphere* **2024**, *15*, 689. https://doi.org/10.3390/atmos15060689

Academic Editors: Lars Hoffmann and Yi Heng

Received: 30 April 2024
Revised: 30 May 2024
Accepted: 4 June 2024
Published: 6 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Weather and climate exert profound effects on daily activities [1–6], human health [7–10], economic prosperity [11–13], ecosystem equilibrium [14,15], and global governance [16–18]. Weather forecasting primarily concerns the prediction and characterization of short-term atmospheric conditions, whereas climate modeling focuses on long-term patterns and alterations in atmospheric parameters. The escalating occurrence of extreme events and significant shifts in climate conditions have escalated the need for more accurate and timely weather and climate forecasting [19–21].

Numerical Weather Prediction (NWP), currently the most advanced method for weather forecasting, predicts future atmospheric states through solving a set of partial differential equations (PDEs) that describe atmospheric motion and evolution from given initial conditions. Physical parameterizations are employed to approximate the effects of small-scale processes like cloud microphysics, turbulence, and radiation that cannot be explicitly resolved within the models. Data assimilation techniques are used to integrate observations with model predictions, enhancing the accuracy of these initial conditions [22]. High-performance computing platforms support these complex calculations, handling the computations over high-resolution grids across global or regional boundaries, and

providing precise and timely predictive services for both weather forecasting and climate modeling [23]. NWP is regarded as the most effective approach, with mainstream meteorological agencies depending on it to deliver weather forecasts [24]. The success of NWP reflects a comprehensive understanding of the atmosphere by meteorologists and represents the pinnacle of interdisciplinary scientific research, addressing complex engineering challenges [25].

However, the further development of NWP models is limited by the extraordinarily high cost of computing [26]. NWP is facing challenges that include but are not limited to errors in initial conditions, biases in numerical models, parameterization issues, and especially the substantial computational costs involved due to continuous resolution increase [24,27]. Concurrently, the proliferation of observation technologies generates hundreds of millions of points of data daily, which need to be processed by forecasting systems. NWP requires a long time to produce a prediction on a supercomputer, with the computational demands of global forecasts rivaling those of simulating the human brain and the early universe [28,29]. As technological advancements approach the post-Moore's Law era, concerns are growing about the future capacity of high-performance computing to sustain improvements [30]. Neural networks, in theory, can potentially capture the nonlinear relationships between variables in any dataset by learning from large amounts of data [31,32]. The potential of neural networks lies in their ability to provide rapid inference results at low computational costs once trained [33–35]. This capability presents a promising alternative to overcoming the computational limitations of NWP methods.

This review outlines the key methods in the development of deterministic forecasting with data-driven models from 2018 until now, structured to mimic the workflow of deep learning, including the selection of datasets, model design, training process, evaluation of the computational acceleration and forecasting effectiveness compared to traditional NWP methods. The rapid evolution of data-driven models has produced impressive results. However, these methods, primarily driven by computer scientists, continue to confront challenges such as physical consistency and accurately quantifying uncertainty, underscoring the need for further insights from meteorological experts. This review will assist meteorologists in understanding the development process of data-driven models and their integration with weather forecasting and climate prediction. It also identifies ongoing challenges and potential directions for future research, suggesting optimizations and facilitating the convergence of computational and meteorological expertise to enhance forecasting outcomes.

2. Evolution of Weather Forecasts

The evolution of weather forecasting has progressed in tandem with humanity's deepening comprehension of the natural world and advancements in science and technology. Prior to the advent of modern civilization, forecasting was predominantly based on the direct observations and personal experiences of scholars. With the dawn of the industrial era, the development and use of meteorological instruments like thermometers, barometers, and hygrometers marked the inception of modern meteorology. This period initiated the systematic recording and statistical analysis of weather data. The advent of telegraphy enabled the rapid communication and exchange of weather information and forecasts across various regions. In the early 20th century, Cleveland Abbe and Vilhelm Bjerknes proposed the idea of predicting future weather changes using partial differential equations governing atmospheric conditions [36,37]. By the mid-twentieth century, the advent of computers revolutionized meteorology by providing the computational power necessary to solve these complex equations, leading to the first successful numerical weather forecast [38]. In the late 20th century, the deployment of weather satellites provided novel data and perspectives on atmospheric and climatic phenomena.

In recent years, the advancement of supercomputing and big data analytics has significantly enhanced the resolution of global numerical weather forecasting systems [39]. Numerous countries now operate their own forecasting centers, delivering cutting-edge

weather forecasting services. Notable examples include the European Centre for Medium-Range Weather Forecasts (ECMWF) in Europe, which is considered one of the most advanced numerical forecasting systems in use [24], the Global Forecast System (GFS) in the United States, and the China Meteorological Administration (CMA).

Artificial Intelligence (AI) for science has shown its potential to tackle complex scientific challenges across various fields like biochemistry [40–42], physics [43], astronomy [44], and materials science [45,46]. Earlier studies have integrated AI methods into numerical forecasting, covering aspects like data processing [47,48], parameterization surrogates [49], and forecast post-processing [50]. Recent research has increasingly applied AI technology to improve weather forecasting in areas such as precipitation [51–55], tropical cyclones [56,57], wind speed [58], and typhoons [59], among other applications [60]. These methods have shown promising potential in improving both prediction accuracy and computational efficiency.

With the development of deep learning techniques, there has been a growing interest in using deep learning models to completely replace numerical models. Since 2018, some research initiatives have adopted a more ambitious approach, employing deep learning techniques for data-driven global weather forecasting [61–64].

3. Data-Driven Models: Methodologies and Performance Evaluation

This chapter will outline the essential steps involved in developing a data-driven forecasting model. Figure 1 illustrates the critical components of such a model. The process begins by defining the problem and desired outcomes. Subsequently, an appropriate dataset and model architecture are selected to align with specific performance goals, budget constraints, and application requirements. Next, the data undergoes preprocessing, and the model is customized and optimized on chosen hardware using acceleration strategies to enhance efficiency and computational performance. The predictive effectiveness of the model is then evaluated. Additionally, Figure 1 highlights strategies to improve model performance and capability, including the optimization of datasets, model structures, training process, and comprehensive evaluations.

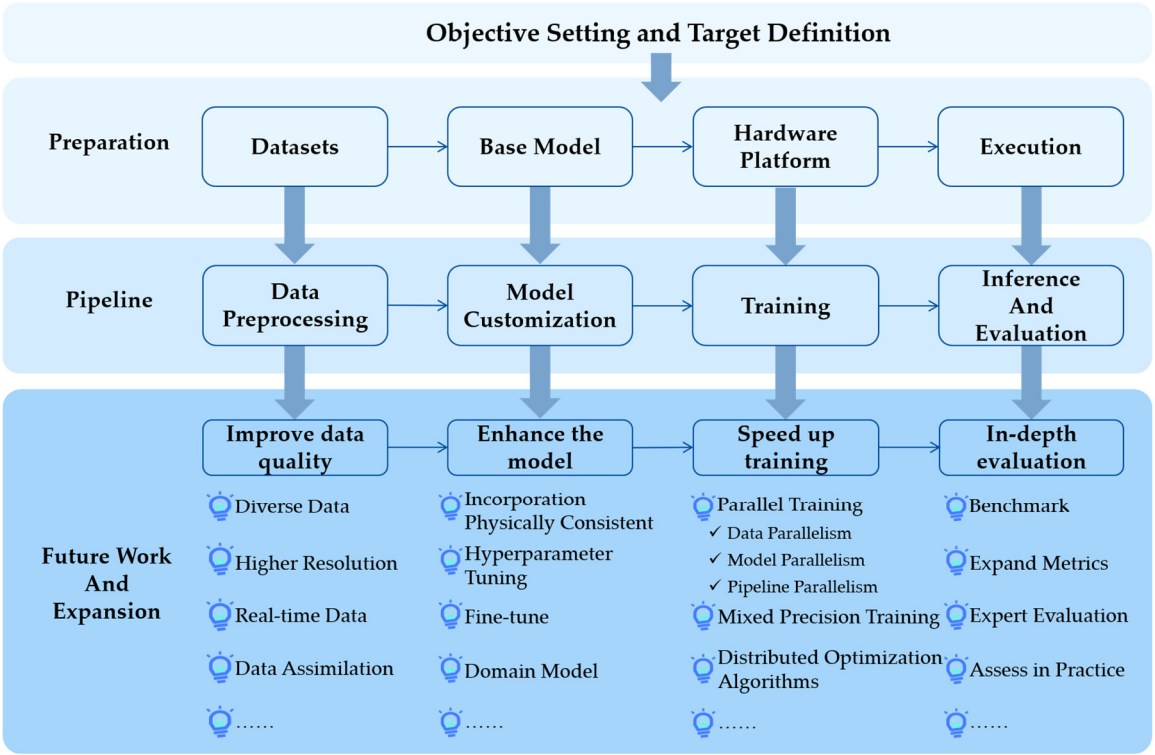


Figure 1. Key steps of data-driven models and future improvement perspectives.

3.1. Datasets

Data serves as the cornerstone for any deep learning model. In deep learning, insights are derived by learning from the features in the training data and adjusting parameters through optimization processes to minimize prediction errors. A comprehensive and robust dataset is crucial for mitigating overfitting and ensuring model reliability. Consistently, research across various fields has underscored the correlation between the quality of data and the efficacy of the model [65–67].

Data-driven models require large volumes of high-quality, diverse data with high spatio-temporal resolution. Observation data, including ground observations, radar data, and satellite imagery, undergo complex processing including data assimilation before being used for training. Well-known datasets that are employed in data-driven models include ERA5 [68], WeatherBench [69,70], climateBench [71], datasets of Coupled Model Intercomparison Project Phase 6 (CMIP6) [72], etc.

ERA5 is the fifth generation of ECMWF atmospheric reanalysis of the global climate, produced by the Copernicus Climate Change Service (C3S). This dataset integrates a vast array of historical observations—including ground observations, satellites, radars, sondes, buoys, aircraft, and ships—using a sophisticated modeling and data assimilation system. It provides detailed information on various atmospheric, land-surface, and sea-state parameters, accompanied by uncertainty estimates. Notably, ERA5 features a high resolution on regular latitude–longitude grids at $0.25^\circ \times 0.25^\circ$, and employs 137 atmospheric layers to analyze the atmosphere from the surface up to an altitude of 80 km. This granularity facilitates a comprehensive understanding of climatic and weather conditions across different scales and elevations. The publicly available dataset currently covers the period from 1979 to August 2019. It is regarded as the best-known estimation for most atmospheric variables, but it is not a perfect reflection of reality [24]. Most models were trained using ERA5 data [24,65,73–77].

WeatherBench, specifically designed for evaluating and benchmarking data-driven weather forecasting models, utilizes data from ERA5 formatted suitably for deep learning networks and other data-driven models, covering the period from 1979 to 2018 [69]. The dataset offers different spatial resolutions, including 5.625°, 2.8125°, and 1.40525°, and features 13 selected vertical layers. Evaluation indicators, such as Root Mean Square Error (RMSE), Anomaly Correlation Coefficient (ACC), and Stable Equitable Error in Probability Space (SEEPS) are provided. An update, WeatherBench2, enhances the original by increasing the spatial resolution to 0.25° and incorporating the Continuous Ranked Probability Score (CRPS) [70].

ClimateBench inspired by WeatherBench, aims to assess and improve data-driven climate prediction models. It utilizes data from the Norwegian Earth System Model, version 2 (NorESM2), with the evaluation index being the normalized root-mean-square error (NRMSE), which measures the models' effectiveness in simulating climate dynamics [71].

The CMIP6 is a global climate model comparison project, and the related datasets contain simulation results of mainstream climate system models and earth system models on temperature, precipitation, and other variables derived from hundreds of climate models. These variables describe various aspects of the Earth system's weather and climate [72].

3.2. Models Adaptations and Training

Over the years, scientists have developed fundamental models that have revolutionized problem-solving in fields such as image processing and natural language understanding. These foundational models form the backbone of rapid advancements in various areas of AI, including weather forecasting and climate modeling. In the subsequent sections, we will examine key existing methods, emphasizing historical advancements and significant outcomes, to thoroughly understand models construction and training. Table 1 lists the basic models used by the various methods.

Table 1. A brief introduction and usage of the basic model.

Basic Model	Description	Method
MLP (Multi-Layer Perceptron)	early neural network suited for nonlinear challenges	Dueben et al. [62]
CNNs (Convolutional Neural Networks)	efficiently processing spatial data and extracting features	Scher et al. [61] Weyn et al. [64] Weyn et al. [78]
ResNet (Residual Network)	enabling deeper networks through efficient residual connections	Rasp et al. [79]
GNN (Graph Neural Network)	capturing spatial and temporal dynamics critical in fluid dynamics	Keisler et al. [80] GraphCast [65]
Transformer	processes input data through self-attention and feedforward layers	FourCastNet [73] FengWu [74] Pangu [24] ClimaX [81] FengWu-GHR [77] Fuxi [75]
EPD (Encode–Process–Decode)	a differentiable model with deep learning	NeuralGCM [76]

3.2.1. Model Based on MLP

The Multi-Layer Perceptron (MLP) [82], an early neural network form, features one or more hidden layers enhancing its ability to tackle a broad array of nonlinear problems, making it suited for meteorological applications. However, the MLP lacks inherent spatial and temporal awareness, treat input features independently, and struggle with handling high-dimensional weather data effectively.

One of the initial efforts to use deep learning models as an alternative to traditional NWP models for medium-term forecasting was pioneered by Dueben et al. [62,83,84]. Faced with challenges inherent in traditional NWP methods, they explored the use of ERA5 data to train a "toy model" model. Although their model was tasked only with predicting a single parameter—the geopotential height at 500 hPa (Z500)—it demonstrated competitive results compared to coarse resolution dynamic models (TL21) in the short term. This method reveals the possibility of data-driven weather forecasting for subsequent research.

3.2.2. Models Based on CNNs

Convolutional Neural Networks (CNNs) [85] utilize convolutional layers to process weather data in a manner that preserves the inherent spatial structure and relationships within the data. Weather data, such as temperature, humidity, and pressure, typically exhibit spatial patterns that extend across geographic regions displayed in a grid-like format. These phenomena display spatial layers and dependencies that vary from small scales to large scales. This enables the model to capture spatial dependencies and automatically extract features from the grid-like topology of the data. However, the primeval CNNs cannot capture time dependencies.

The CNN model network trained by Scher et al. [61], which incorporated four atmospheric variables, was capable of successfully simulating and predicting the state of the general atmospheric circulation model. Weyn et al. [64] used only two atmospheric variables—500-hPa geopotential height and 700–300-hPa thickness—and by integrating training with both CNN and LSTM (Long Short-Term Memory, known for its ability to learn dependencies in long time series [86]) Weyn et al. employed a U-Net (a U-shaped convolutional neural network [87]) architecture-based CNN, which performed data mapping on a cubed-sphere grid to handle global data. This model was able to provide forecasts of several crucial atmospheric variables with a lead time of up to 7 days and maintain stability on an annual timescale [78]. These pioneering approaches, despite not rivaling

state-of-the-art NWP accuracy, demonstrate deep learning's potential, as recognized by Schultz et al. in 2021 who noted the field's early stage in meteorology [88].

3.2.3. Model Based on ResNet

Residual Networks (ResNets) [89] improve weather forecasting by enabling deeper networks to be trained without the performance decline usually seen with increasing depth. This is achieved through residual connections that help maintain performance levels. ResNets are particularly effective for handling complex spatial and temporal patterns in weather data, as they ensure that both low-level and high-level features are effectively integrated into the decision-making process. But it can not inherently handle temporal dependencies as effectively. Rasp et al. applied ResNet to medium-term weather forecasting, expanding the number of predictors to 11 atmospheric variables across seven vertical levels, albeit at a lower spatial resolution of 5.625°. Their approach achieved superior results on the WeatherBench dataset compared to previous methods [69,79].

3.2.4. Models Based on GNN

Methods that use latitude and longitude grids struggle with varying space sizes and detailed features at the poles. The Graph Neural Network (GNN) [90] effectively processes weather data by converting it into an irregular graph structure, where nodes represent entities and edges define their relationships. GNNs capture local and global contexts by exchanging messages between nodes, allowing for the modeling of complex interactions in an adaptable structure that better handles irregular spatial resolutions and overcomes limitations of traditional latitude–longitude grids [84,91–93]. However, GNN structures struggle with high computational demands for training and have a strong dependence on the quality of graph construction.

Keisler et al.'s method utilizes a GNN architecture, consisting of an Encoder, Processor, and Decoder, to map raw physical data such as temperature and wind speed to an intermediate feature space on an icosahedron grid. The model is trained with multi-resolution data and employs a multi-step loss function to adjust the model parameters. After training, the processed data are mapped back to the original data space [80].

The GraphCast method uses a "multimesh" structure processed by GNN. It creates a graph with high spatial resolution by iteratively refining a regular icosahedron. This "multimesh" structure allows GraphCast to capture much longer spatial interactions than traditional NWP methods [65].

3.2.5. Models Based on Transformer

The Transformer model's robust attention mechanism has shown significant potential in data-driven weather forecasting [94], effectively managing wide-ranging weather data dependencies with parallel processing. Vision Transformer (ViT) [95] and Swin Transformer [96] improve Transformer applicability to image tasks. ViT segments images into fixed blocks for linear embedding and global attention, while Swin Transformer uses Window-based Multi-head Self-attention (W-MSA) for local window computations, offering linear complexity with the number of patches (O(N)). Swin Transformer V2 [97] optimizes this architecture, enhancing its performance and robustness.

The FourCastNet [73] model utilizes a combined architecture of the Adaptive Fourier Neural Operator (AFNO) [98] and ViT model. The AFNO module facilitates the processing of high-resolution data, while the ViT efficiently manages long-range dependencies, all while maintaining computational efficiency. It operates on a grid of tokens derived from input frames and refines these through a series of Transformer layers. The model is trained in phases, including a pre-training stage and a fine-tuning stage, to predict atmospheric conditions at subsequent time steps.

Pangu [24] utilizes a Patch Embedding technique to efficiently process three-dimensional weather data, focusing on essential features while simplifying spatial resolution. It integrates this data into a deep network, using a unique 3D Earth-specific Transformer

architecture enhanced by an Earth-specific positional bias to accurately track atmospheric dependencies, employing a hierarchical temporal aggregation strategy for training and forecasting at 1 h, 3 h, 6 h, and 24 h intervals.

FengWu [74] treats the high-dimensional weather data as distinct modalities, feeding them into modality-specific encoders. It simulates the interactions among all atmosphere variables through a Transformer-based fusion. Finally, the forecast results are derived separately after the cross-modal Transformer with modal-specific decoders. To improve forecast lead time and reduce error accumulation, a Replay Buffer mechanism is introduced.

The FuXi [75] model revolves around the U-Transformer, assembled from 48 repeated Swin Transformer V2 blocks. This architecture employs a scaled cosine attention mechanism to effectively process the complex, high-dimensional weather data. This model comprises three components: FuXi-Short, FuXi-Medium, and FuXi-Long, designed to enhance the accuracy of long-term weather forecasting and mitigate the accumulation of forecast errors.

ClimaX [81] is a foundational model built on ViT, trained using data from the CMIP6, that can work with any forecast lead time. It introduces a variable tokenization mechanism that segments each variable's spatial map into patches and embeds each patch, allowing the model to handle a varying number of variables flexibly. It incorporates a variable aggregation module that performs cross-attention operations at each spatial location, reducing sequence length and computational cost through the queries of cross-attention.

The FengWu-GHR [77] aims to deliver highly accurate weather forecasts for up to 10 days at an ultra-high resolution of 0.09°. Initially, the model was pre-trained using data at a 0.25° resolution to preserve the physical information from the original dataset. These data were then up-sampled to a 0.09° resolution using the Spatial Identical Mapping Extrapolate (SIME) method. The model incorporates three key components: a 2D patch embedding layer, stacked Transformer blocks, and a deconvolution layer, forming a meta-model based on the ViT.

3.2.6. Hybrid Model with Physical Constraints

ACE (AI2ClimateEmulator) [99] employs the SFNO [100] architecture, which uses spherical harmonic transforms to enable global convolutions while respecting the symmetries of the spherical domain. This enhances the model's ability to capture atmospheric dynamics and maintain physical consistency. ACE generates training data using the FV3GFS model, which is the atmospheric component of the United States weather model. The SFNO architecture allows the model to capture large-scale atmospheric patterns and interactions efficiently.

NeuralGCM combines the strengths of both data-driven models and physics-based dynamical models. It maintains a differentiable dynamic core that solves discretized Navier–Stokes equations [101]. The model interfaces with data using the Sigma coordinate system, which enriches its understanding of atmospheric conditions with features such as spatial derivatives, surface types, and topography. It utilizes the Encode–Process–Decode (EPD) architecture to construct a deep learning framework. The Encoder maps input data to an intermediate latent representation space, the process network parameterizes the complex physical processes of the atmosphere, and the Decoder translates these outputs into weather and climate forecasts. By incorporating neural networks, it leverages their ability to learn complex patterns from data while maintaining the physical consistency of atmospheric dynamics. While providing computational efficiencies compared to NWP, it still demands more resources than pure deep learning models.

3.3. Evaluation

3.3.1. The Speed Benefits of Data-Driven Models

The HRES, a high-resolution configuration of the Integrated Forecasting System (IFS) utilized by the ECMWF, spends one hour to produce a 10-day weather forecast at a detailed latitude and longitude resolution of 0.1 degrees [65]. Similarly, the IFS needs about an hour and a half on 1530 CrayXC40 computer nodes to complete a 15-day ensemble forecast with

a resolution of 18 km [102]. In developed countries, it is common to use supercomputers for detailed weather predictions. However, in developing regions like parts of Africa, the lack of advanced computing resources limits access to accurate weather forecasts, impacting the safety and protection of lives and property [103].

Deep learning models are highly parallel and need substantial computing resources for training. To meet this demand, several companies have created GPUs and TPUs specifically tailored to speed up the training process for these models [104–106]. These devices are optimized for deep learning tasks, featuring Tensor Cores specifically engineered for such applications, equipped with high-bandwidth memory, and supported by a comprehensive software ecosystem. Most current weather forecasting models are trained using accelerated technologies provided by companies like Nvidia and Google. Table 2 illustrates the performance of the most used hardware.

Table 2. Performance of the widely used GPUs.

Hardware	Performance (TFLOPS)	MEMORY (GB)	Memory Bandwidth
Cloud TPU v4	275	32	1200 GB/s
NVIDIA A100 GPU	312~1248	40	1.6 TB/s
NVIDIA V100 GPU	112~125	32/16	900 GB/s

Both GPUs and TPUs offer significant benefits for deep learning, but they differ in terms of design, performance, versatility, and software support. GPUs are versatile, while TPUs excel in performance and efficiency for tensor operations in deep learning. Although training a deep learning model can take days to weeks, once trained, it can produce predictions up to magnitudes faster than traditional methods like the IFS, delivering both high accuracy and efficiency [24,73]. Table 3 provides information on the training duration and inference times for these models.

Table 3. Details the time required to train the weather forecasting model and the time it takes to perform inference, as mentioned in the original papers.

Method.	Hardware	Training Cost	Inference Cost
Weyn et al. [78]	1 NVIDIA V100 GPU	2–3 days	4-week forecast in less than 0.2 s on one GPU
Keisler et al. [80]	1 NVIDIA A100 GPU	5.5 days	5-day forecast takes about 0.8 s on one GPU
FourCastNet [73]	64 NVIDIA A100 GPU	16 h	1 week-long forecast in less than 2 s on one GPU
Pangu [24]	192 NVIDIA V100 GPU	16 days	5-day forecast takes 1.4 s on one GPU
FengWu [74]	32 NVIDIA A100 GPU	17 days	10-day forecast takes 0.6 s for on one GPU
GraphCast [65]	32 Cloud TPU v4	about 4 weeks	10-day forecast takes less than 1 min on one TPU
ACE [99]	4 NVIDIA A100 GPU	63 h	1 day simulate takes 1 s on one GPU
NeuralGCM [76]	16~256 Cloud TPU v4	1 day to 3 weeks	10-day forecast takes from 2.5 s to 119 s in different spatial resolutions on one TPU

3.3.2. Evaluating Forecast Quality

Current data-driven models mostly use RSME and ACC as performance measures and perform as well as or better than the top methods like IFS and HERS. This performance varies depending on the forecast time, the number of variables considered, and the spatial and temporal resolutions employed. However, for practical purposes, it is essential to evaluate these models comprehensively. The specific parameter details of these methods are illustrated in Table 4.

Weather forecasting can generally be categorized into several types based on the timeframe: nowcasting, short-term, medium-range, and long-range or extended range forecasts. Each type focuses on different aspects and involves different prevention strategies. Due to the nature of weather systems, errors tend to accumulate over time, leading to a sharp drop in accuracy for longer forecasts [74,75,107].

Table 4. Details the time required to train the weather forecasting model and the time it takes to perform inference, as mentioned in the original papers.

Method	Forecast Time (Days)	Atmospheric Variables	Vertical Pressure Levels	Surface Variables	Number of Total Variables	Spatial Resolution (°)
Pangu [24]	7	5	13	4	69	0.25
GraphCast [65]	10	6	37	5	227	0.25
FengWu [74]	11.5	5	37	4	189	2.25
Fuxi [75]	14.5	5	13	5	70	0.25
Fengwu-GHR [77]	10	5	13	4	69	0.09
NeuralGCM [76]	15	7	32	0	224	0.7, 1.4, 2.8

Current research primarily concentrates on enhancing the accuracy of medium-range weather forecasting. Pangu stands out as the first model to surpass the IFS in accuracy for 7-day medium-range forecasts. Its performance has been proven to be superior to that of the ECMWF's IFS across all test variables. GraphCast extends its forecasting advantage to 10 days. This model trained on 227 variables, significantly outperformed ECMWF-HRES on 90% of the 1380 validation targets, excelled more than Pangu on 99.2% of the 252 targets. This performance edge starts to manifest from 3.5 days onward and increases progressively [24]. GraphCast also benefits from being retrained periodically with recent data, demonstrating that updates with the latest data can further enhance its predictive performance [65]. Using ACC greater than 0.6 as the criterion for a skillful weather forecasting system, FengWu was the first to extend the skillful forecast lead time to 10.75 days for the Z500 (Geopotential Height at 500 hPa) variable and to 11.5 days for the T2M (2-Meter Temperature) variable. When compared directly to GraphCast, FengWu performed better on 80% of the 880 targets. While FengWu showed a similar performance to GraphCast in forecasts ranging from 1 to 5 days, it surpassed GraphCast in accuracy for forecasts between 5 and 10 days [74]. FuXi has extended the effective ACC, maintaining a value greater than 0.6, for the Z500 variable up to 10.5 days and for the T2M variable up to 14.5 days. This marks an achievement where FuXi surpasses FengWu in the T2M forecast by 3 days. When compared with GraphCast, both models outperform ECMWF-HRES in the 10-day forecast range. In the first 7 days, FuXi and GraphCast show nearly identical performance levels significantly better than ECMWF-HRES. With the extension of the forecast time, FuXi showed better performance, especially in more than 7 days; FuXi gradually significantly outperformed GraphCast. However, it is noted that the ACC of both FuXi and GraphCast begins to decline significantly after 5 days, highlighting a challenge in maintaining forecast accuracy over longer periods. NeuralGCM is a physically consistent model capable of handling both weather prediction and climate modeling. It was trained at three different resolutions—2.8°, 1.4°, and 0.7°—to balance computational cost and accuracy. NeuralGCM-0.7° outperforms ECMWF-HRES in the 10-day forecast. In the 3-day forecast, NeuralGCM-0.7° and GraphCast perform equally well, achieving the best results [76].

The FengWu-GHR model not only improves prediction accuracy and lead time but also triples the spatial resolution for its 10-day forecast operating at a 0.09° horizontal resolution, a first in global weather forecasting. FengWu-GHR's performance matches or exceeds ECMWF-HRES in the 10-day forecast at this finer resolution. However, it is noteworthy that ECMWF-HRES's ACC on parameters like Z500 and T500 (Temperature at 500 hPa) on the 10th day falls below 0.6 at this resolution.

For extending the forecast time to the long-term climate scale, ClimaX [81] can generate forecasts from a few hours to weeks or even months at multiple resolutions (5.625° and 1.40625°). Over longer forecast times, such as 2 weeks and 1 month, ClimaX's performance is comparable or superior to IFS. In climate projection and climate model downscaling tasks, ClimaX's performance matches the best baseline methods. ACE [99] enables the evaluation of physical laws like mass and moisture conservation and remains stable for 100 years on a comprehensive 100 km resolution. The model faithfully replicates the climate

of the reference model for over 90% of tracked variables, with 100× reduced computational time and 100× more energy efficient compared to the reference model. NeuralGCM shows significant capabilities in both medium-range weather forecasting and long-term climate simulations. In two-year simulation tests, it accurately reproduces seasonal cycles, tropical cyclones, and monsoon systems [81]. Furthermore, NeuralGCM can operate stably for extended periods ranging from years to decades, without numerical instability or climate drift. Overall, it performs comparably or better than some of the best existing models in terms of accuracy, physical consistency, effective management of the water budget, and the ability to simulate climate across multiple tasks and dimensions [76]. Figure 2 illustrates the key approaches in the development of data-driven models from 2018 to the present.

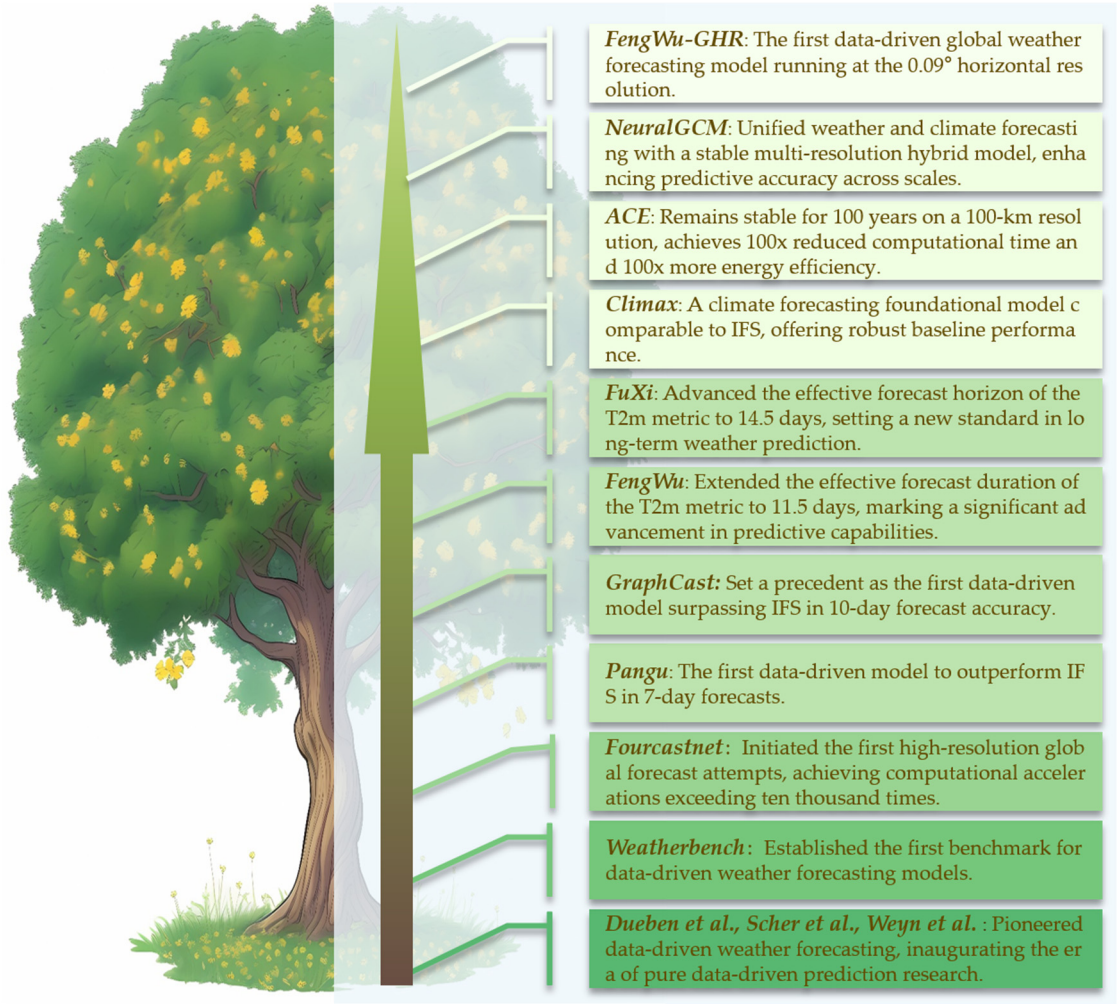


Figure 2. Key methods emerging in the development of data-driven models and their decisive meanings [61,62,78].

3.3.3. Assessing Ensemble Forecasting Capabilities

The efficiency of the data-driven method in quickly generating forecasts enables the integration of perturbations to produce ensemble forecasts effectively. By adding random perturbations to the initial state and averaging the results, Pangu's ensemble approach typically yields slightly less accurate results than the single-member method for short-term forecasts. However, for medium-range forecasts spanning 5 to 7 days, it proves to be more accurate than the single-member forecasts [24]. FuXi introduces perturbations both in initial conditions and model parameters. This strategy enhances the forecast's ability to estimate uncertainties over an extended period, up to 15 days. Its performance matching the ECMWF's ensemble average highlights its robustness in longer-

term forecasting scenarios [75]. NeuralGCM-ENS, an enhanced version of NeuralGCM, incorporates ten additional space-time correlated Gaussian Random Fields (GRF). This redesign has allowed it to achieve performance comparable to the ECMWF ensemble forecasts over a 1–15-day period [76]. Integrating these methods could harness their strengths across different forecast periods and uncertainties, potentially leading to a robust ensemble forecasting system that improves predictive accuracy and reliability.

4. Opportunities and Challenges

4.1. Advantages

Traditional NWP models built with empirical equations and parametric schemes are updated slowly and may not adapt quickly to the Earth's rapidly changing climate influenced by human activities. In contrast, data-driven models utilize large historical datasets, allowing deep learning networks to capture complex climate interactions. These trained models can process real-time data streams quickly, enabling near real-time predictions on standard computing devices and outperforming NWP methods in generating ensemble predictions. By continuously updating and iterating models, data-driven approaches offer efficient data utilization, automatic feature extraction, real-time forecasting, and cost-effective iteration.

4.2. Limitations

Despite transforming the field with new perspectives and capabilities, developing data-driven weather models presents several challenges.

4.2.1. Weak Interpretability

Deep learning models are often viewed as 'black boxes', lacking transparency. This is a major criticism that comes from scientific tasks like weather forecasting and climate prediction as interpretability is crucial for enhancing model credibility, improving decision support, refining methodologies, facilitating learning, and deepening theoretical understanding. Predictability in weather forecasting provides vital information about risks and uncertainties essential for decision-making and helps scientists gain a deeper understanding of meteorological phenomena. McGovern et al.'s work examined the interpretability of early machine learning models in some meteorological applications [108]. Liu et al. conducted research on the explainability of a post-processing model based on deep learning [109]. Yet, no studies have currently focused on the interpretability of data-driven global weather forecasting models.

4.2.2. High Reliance on High-Quality Training Data

Currently, data-driven models that show excellent results typically use datasets generated using NWP methods, which limits the model capability and influences the learning effectiveness and generalization ability. The enhancement of the models' predictive capabilities also depends on the density of information within the dataset, such as the number of variables, vertical levels, and resolution. Data assimilation merges observational data from diverse sources into a unified framework, providing a detailed representation of the atmospheric state. However, the precision of current data assimilation methods in managing high-resolution and high-dimensional data structures remains a challenge.

4.2.3. Uncertainty Quantification

Deep learning models, which inherently contain uncertainties, have not been proven to quantify the uncertainty of the physical world more reasonably. A key issue is that the current models easily lead to error accumulation that can be magnified during continuous inference. Studies have shown that performing multiple rollouts with shorter forecasting intervals results in greater errors compared to fewer rollouts with longer forecasting intervals [24]. The "hierarchical temporal aggregation strategy" proposed by Pangu [24], the "Replay Buffer mechanism" proposed by FengWu [74], and the "cascade innovation

mechanism" proposed by FuXi [75], have all proven effective in reducing error accumulation. Establishing and applying a comprehensive integral stability theory for data-driven models is crucial for improving their applicability and robustness across different scenarios. Additionally, balancing the forecast interval and associated errors remains a challenge. Addressing this issue will require rigorous mathematical analysis and extensive testing to ensure model stability under various conditions.

4.2.4. Unsatisfactory in Extreme Cases

If the dataset lacks a comprehensive description of extreme conditions, the generalizability of data-driven models to extreme weather scenarios is limited. Deep learning typically produces smooth results. Although deterministic forecasting methods like Pangu [24], GraphCast [65], and NeuralGCM [76] demonstrate promising results in specific extreme cases—achieving s prediction accuracy on par with or better than the advanced HRES—they still tend to underestimate the magnitude of extreme events. Moreover, these methods do not systematically evaluate the model's capacity to manage extreme situations.

4.2.5. Incomplete Evaluation

Current evaluation methods primarily rely on metrics such as RMSE, ACC, and CRPS, which may not fully capture the model's effectiveness. Often, these data-driven models are assessed by data scientists lacking expertise in weather and climate evaluation, leading to flawed or incomplete assessments [110]. Therefore, it is crucial to establish a standard, comprehensive, easy-to-use, and widely recognized evaluation benchmark to ensure the reliability and applicability of these methods [69–71]. Although studies have validated Pangu's effectiveness in an operational-like context [111], these data-driven models still require further verification. In this area, NeuralGCM [76] has established a good foundation.

4.3. Future Research

Data-driven models are still evolving, showing that substantial progress is needed to significantly advance the fields of meteorology and climatology. A promising direction involves new data assimilation methods that offer more precise global search capabilities and detailed uncertainty analyses, which are crucial for understanding the reliability of predictions. With the rapid advancement of remote sensing and other observational technologies, the volume of meteorological data has dramatically increased. It is essential for new data assimilation methods to efficiently process these vast datasets, extracting vital information while minimizing computational resource consumption. Some practical, application-oriented studies focus on training models directly using real data. Conformer, for example, uses data from weather stations to predict temperatures and wind speeds [112]. These studies could enlighten the practical application of data-driven models that train models with actual data and assess their performance based on real atmospheric conditions.

Integrating deep learning with traditional methods can potentially lead to the creation of comprehensive models that maintain physical rationality and integral stability. For example, the NeuralGCM model demonstrates the importance of incorporating physical constraints appropriately for long-term stable climate forecasts. Data-driven models need to improve the dynamic adaptability and accuracy of predictions by learning from both historical and real-time data, ensuring physical rationality and system stability. The key to successful integration lies in developing data-driven models that adhere to physical laws, fully utilize data, and maintain integral stability.

The development of large language models (LLMs) and generative AI technologies provides new opportunities for creating high-accuracy simulations and forecasts that can quickly adapt to different scenarios. The rapid advancement of LLMs has continuously shown that larger models tend to yield better performance [113]. For instance, the GPT-3 model has 175 billion parameters, whereas the most parameter-rich forecasting model,

FengWu-GHR, has 4 billion parameters [77,114]. However, larger models pose greater challenges in training, and methods from LLM research could inform training optimization.

Notably, the challenges tackled by natural language processing differ from those in weather forecasting and climate modeling, which calls for further research to explore how the number of parameters affects forecasting accuracy. Moreover, training deep learning models involves numerous hyperparameters that depend on extensive expert knowledge; tuning hyperparameters also requires expertise based on prior knowledge, experience, and trial and error. Automation in hyperparameter tuning can significantly reduce the time and expertise required to find optimal settings [115,116]. Automated tools like Hyperopt, Optuna, and Google's Vizier can assist meteorologists and data scientists by efficiently exploring parameter spaces, but expert oversight remains essential to guide the tuning process and interpret results effectively [117–119].

From a broader perspective, the underlying models used in quite a few data-driven models are adapted from structures that were originally developed for other domains, not specifically for weather forecasting and climate modeling. While these models perform well in reproducing known patterns, this can become a bottleneck. To overcome this limitation, the unique characteristics of weather data must be explored, and foundational models that are specifically tailored for weather data should be designed. This task is highly challenging but essential for advancing the field.

5. Summary

This literature review examines the evolution of data-driven weather forecasting and climate models. NWP encounters challenges such as inherent atmospheric uncertainties and high computational costs, slowing down the NWP's growth in predictive capability.

Since 2018, neural network models have been proposed as alternatives to traditional NWP methods. This paper reviews the key developments in this area, detailing the workflow for developing deep learning models, which includes training dataset selection, model architecture design, training process, computational acceleration, and prediction effectiveness. The paper highlights the potential of data-driven models to generate forecasts significantly faster than traditional numerical methods. Notably, GraphCast has demonstrated superior performance in 10-day forecasts, Fuxi has extended the effective prediction time to 15 days, NeuralGCM has explored the fusion of numerical models and neural networks to balance computational speed with physical consistency, and FengwuGHR has enhanced the resolution of 10-day forecasts to 0.09°.

However, challenges persist, including difficulties in quantifying uncertainty, poor model interpretability, lack of physical consistency, underestimation of extreme cases, relatively simple evaluation methods, and untested practical applications. Future research directions involve developing larger, more complex models, integrating more physical constraints, and refining evaluation frameworks to improve the performance and reliability of data-driven weather forecasting and climate models. Overall, this review underscores the interdisciplinary nature of weather forecasting and climate modeling, emphasizing the need for collaboration between meteorologists and computer scientists to optimize forecasting processes and address emerging challenges.

Author Contributions: Writing—original draft preparation, Y.W.; participation in the discussion, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No.U2242210).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

ACC	Anomaly Correlation Coefficient
ACE	AI2ClimateEmulator
AFNO	Adaptive Fourier Neural Operator
AI	Artificial Intelligence
C3S	Copernicus Climate Change Service
CMA	China Meteorological Administration
CMIP6	Coupled Model Intercomparison Project Phase 6
CNNs	Convolutional Neural Networks
CRPS	Continuous Ranked Probability Score
ECMWF	European Centre for Medium-Range Weather Forecasts
HRES	High-Resolution Configuration of IFS
EPD	Encode–Process–Decode
ERA5	Fifth Generation Of Ecmwf Atmospheric Reanalysis of The Global Climate
FV3GFS	the Atmospheric Component of the United States Weather Model
GFS	Global Forecast System
GNN	Graph Neural Network
GRF	Gaussian Random Fields
IFS	Integrated Forecasting System
LLMs	Large Language Models
LSTM	Long Short-Term Memory
MLP	Multi-Layer Perceptron
NorESM2	Norwegian Earth System Model, Version 2
NRMSE	Normalized Root-Mean-Square Error
NWP	Numerical Weather Prediction
PDEs	Partial Differential Equations
ResNet	Utilizing Residual Network
RMSE	Root Mean Square Error
SEEPS	Stable Equitable Error in Probability Space
SIME	Spatial Identical Mapping Extrapolate
T2M	2-Meter Temperature
T500	Temperature at 500 hPa
ViT	Vision Transformer
W-MSA	Window-Based Multi-Head Self-Attention
Z500	Geopotential Height at 500 hPa

References

1. Brum-Bastos, V.S.; Long, J.A.; Demšar, U. Weather Effects on Human Mobility: A Study Using Multi-Channel Sequence Analysis. *Comput. Environ. Urban Syst.* **2018**, *71*, 131–152. [CrossRef]
2. Wang, T.; Qu, Z.; Yang, Z.; Nichol, T.; Clarke, G.; Ge, Y.-E. Climate Change Research on Transportation Systems: Climate Risks, Adaptation and Planning. *Transp. Res. Part D Transp. Environ.* **2020**, *88*, 102553. [CrossRef]
3. Palin, E.J.; Stipanovic Oslakovic, I.; Gavin, K.; Quinn, A. Implications of Climate Change for Railway Infrastructure. *WIREs Clim. Chang.* **2021**, *12*, e728. [CrossRef]
4. Bernard, P.; Chevance, G.; Kingsbury, C.; Baillot, A.; Romain, A.-J.; Molinier, V.; Gadais, T.; Dancause, K.N. Climate Change, Physical Activity and Sport: A Systematic Review. *Sports Med.* **2021**, *51*, 1041–1059. [CrossRef]
5. Parolini, G. Weather, Climate, and Agriculture: Historical Contributions and Perspectives from Agricultural Meteorology. *WIREs Clim. Chang.* **2022**, *13*, e766. [CrossRef]
6. Falloon, P.; Bebber, D.P.; Dalin, C.; Ingram, J.; Mitchell, D.; Hartley, T.N.; Johnes, P.J.; Newbold, T.; Challinor, A.J.; Finch, J.; et al. What Do Changing Weather and Climate Shocks and Stresses Mean for the UK Food System? *Environ. Res. Lett.* **2022**, *17*, 051001. [CrossRef]
7. Kim, K.-H.; Kabir, E.; Ara Jahan, S. A Review of the Consequences of Global Climate Change on Human Health. *J. Environ. Sci. Health Part C* **2014**, *32*, 299–318. [CrossRef]
8. Meierrieks, D. Weather Shocks, Climate Change and Human Health. *World Dev.* **2021**, *138*, 105228. [CrossRef]
9. Campbell-Lendrum, D.; Neville, T.; Schweizer, C.; Neira, M. Climate Change and Health: Three Grand Challenges. *Nat. Med.* **2023**, *29*, 1631–1638. [CrossRef]
10. Liu, F.; Chang-Richards, A.; Wang, K.I.-K.; Dirks, K.N. Effects of Climate Change on Health and Wellbeing: A Systematic Review. *Sustain. Dev.* **2023**, *31*, 2067–2090. [CrossRef]
11. Carleton, T.A.; Hsiang, S.M. Social and Economic Impacts of Climate. *Science* **2016**, *353*, aad9837. [CrossRef] [PubMed]

12. Lenton, T.M.; Xu, C.; Abrams, J.F.; Ghadiali, A.; Loriani, S.; Sakschewski, B.; Zimm, C.; Ebi, K.L.; Dunn, R.R.; Svenning, J.-C.; et al. Quantifying the Human Cost of Global Warming. *Nat. Sustain.* **2023**, *6*, 1237–1247. [CrossRef]
13. Malpede, M.; Percoco, M. Climate, Desertification, and Local Human Development: Evidence from 1564 Regions around the World. *Ann. Reg. Sci.* **2024**, *72*, 377–405. [CrossRef]
14. Handmer, J.; Honda, Y.; Kundzewicz, Z.W.; Arnell, N.; Benito, G.; Hatfield, J.; Mohamed, I.F.; Peduzzi, P.; Wu, S.; Sherstyukov, B. Changes in Impacts of Climate Extremes: Human Systems and Ecosystems. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation Special Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2012; pp. 231–290. [CrossRef]
15. Zhao, C.; Yang, Y.; Fan, H.; Huang, J.; Fu, Y.; Zhang, X.; Kang, S.; Cong, Z.; Letu, H.; Menenti, M. Aerosol Characteristics and Impacts on Weather and Climate over the Tibetan Plateau. *Natl. Sci. Rev.* **2020**, *7*, 492–495. [CrossRef] [PubMed]
16. Griggs, D.; Stafford-Smith, M.; Warrilow, D.; Street, R.; Vera, C.; Scobie, M.; Sokona, Y. Use of Weather and Climate Information Essential for SDG Implementation. *Nat. Rev. Earth Environ.* **2021**, *2*, 2–4. [CrossRef] [PubMed]
17. Wilkens, J.; Datchoua-Tirvaudey, A.R.C. Researching Climate Justice: A Decolonial Approach to Global Climate Governance. *Int. Aff.* **2022**, *98*, 125–143. [CrossRef]
18. Chen, Y.; Zhang, D.; Wu, F.; Ji, Q. Climate Risks and Foreign Direct Investment in Developing Countries: The Role of National Governance. *Sustain. Sci.* **2022**, *17*, 1723–1740. [CrossRef]
19. Stott, P. How Climate Change Affects Extreme Weather Events. *Science* **2016**, *352*, 1517–1518. [CrossRef] [PubMed]
20. Zittis, G.; Almazroui, M.; Alpert, P.; Ciais, P.; Cramer, W.; Dahdal, Y.; Fnais, M.; Francis, D.; Hadjinicolaou, P.; Howari, F.; et al. Climate Change and Weather Extremes in the Eastern Mediterranean and Middle East. *Rev. Geophys.* **2022**, *60*, e2021RG000762. [CrossRef]
21. Brunet, G.; Parsons, D.B.; Ivanov, D.; Lee, B.; Bauer, P.; Bernier, N.B.; Bouchet, V.; Brown, A.; Busalacchi, A.; Flatter, G.C.; et al. Advancing Weather and Climate Forecasting for Our Changing World. *Bull. Am. Meteorol. Soc.* **2023**, *104*, E909–E927. [CrossRef]
22. Pu, Z.; Kalnay, E. Numerical Weather Prediction Basics: Models, Numerical Methods, and Data Assimilation. In *Handbook of Hydrometeorological Ensemble Forecasting*; Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H.L., Schaake, J.C., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–31, ISBN 978-3-642-40457-3.
23. Michalakes, J. HPC for Weather Forecasting. In *Parallel Algorithms in Computational Science and Engineering*; Grama, A., Sameh, A.H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 297–323, ISBN 978-3-030-43736-7.
24. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate Medium-Range Global Weather Forecasting with 3D Neural Networks. *Nature* **2023**, *619*, 533–538. [CrossRef] [PubMed]
25. Tufek, A.; Aktas, M.S. A Systematic Literature Review on Numerical Weather Prediction Models and Provenance Data. In *Proceedings of the Computational Science and Its Applications—ICCSA 2022 Workshops*; Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 616–627.
26. Brotzge, J.A.; Berchoff, D.; Carlis, D.L.; Carr, F.H.; Carr, R.H.; Gerth, J.J.; Gross, B.D.; Hamill, T.M.; Haupt, S.E.; Jacobs, N.; et al. Challenges and Opportunities in Numerical Weather Prediction. *Bull. Am. Meteorol. Soc.* **2023**, *104*, E698–E705. [CrossRef]
27. Peng, X.; Che, Y.; Chang, J. A Novel Approach to Improve Numerical Weather Prediction Skills by Using Anomaly Integration and Historical Data. *J. Geophys. Res. Atmos.* **2013**, *118*, 8814–8826. [CrossRef]
28. Bauer, P.; Thorpe, A.; Brunet, G. The Quiet Revolution of Numerical Weather Prediction. *Nature* **2015**, *525*, 47–55. [CrossRef] [PubMed]
29. Govett, M.; Bah, B.; Bauer, P.; Berod, D.; Bouchet, V.; Corti, S.; Davis, C.; Duan, Y.; Graham, T.; Honda, Y.; et al. Exascale Computing and Data Handling: Challenges and Opportunities for Weather and Climate Prediction. *Bull. Am. Meteorol. Soc.* 2024; *published online ahead of print*. [CrossRef]
30. Balaji, V. Climbing down Charney's Ladder: Machine Learning and the Post-Dennard Era of Computational Climate Science. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2021**, *379*, 20200085. [CrossRef] [PubMed]
31. Hornik, K.; Stinchcombe, M.; White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
32. Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G.E. Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. *Nat. Mach. Intell.* **2021**, *3*, 218–229. [CrossRef]
33. Aminabadi, R.Y.; Rajbhandari, S.; Awan, A.A.; Li, C.; Li, D.; Zheng, E.; Ruwase, O.; Smith, S.; Zhang, M.; Rasley, J.; et al. DeepSpeed- Inference: Enabling Efficient Inference of Transformer Models at Unprecedented Scale. In Proceedings of the SC22: International Conference for High Performance Computing, Networking, Storage and Analysis, Dallas, TX, USA, 13–18 November 2022; pp. 1–15.
34. Shuvo, M.M.H.; Islam, S.K.; Cheng, J.; Morshed, B.I. Efficient Acceleration of Deep Learning Inference on Resource-Constrained Edge Devices: A Review. *Proc. IEEE* **2023**, *111*, 42–91. [CrossRef]
35. Menghani, G. Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. *ACM Comput. Surv.* **2023**, *55*, 1–37. [CrossRef]
36. Abbe, C. The physical basis of long-range weather forecasts. *Mon. Wea. Rev.* **1901**, *29*, 551–561. [CrossRef]
37. Bjerknes, V. Das Problem Der Wettervorhersage, Betrachtet Vom Standpunkte Der Mechanik Und Der Physik. *Meteor. Z.* **1904**, *21*, 1–7.
38. Lynch, P. The Origins of Computer Weather Prediction and Climate Modeling. *J. Comput. Phys.* **2008**, *227*, 3431–3444. [CrossRef]

39. Mass, C. The Uncoordinated Giant II: Why U.S. Operational Numerical Weather Prediction Is Still Lagging and How to Fix It. *Bull. Am. Meteorol. Soc.* **2023**, *104*, E851–E871. [CrossRef]
40. Gomes, B.; Ashley, E.A. Artificial Intelligence in Molecular Medicine. *N. Engl. J. Med.* **2023**, *388*, 2456–2465. [CrossRef] [PubMed]
41. Mullowney, M.W.; Duncan, K.R.; Elsayed, S.S.; Garg, N.; van der Hooft, J.J.J.; Martin, N.I.; Meijer, D.; Terlouw, B.R.; Biermann, F.; Blin, K.; et al. Artificial Intelligence for Natural Product Drug Discovery. *Nat. Rev. Drug Discov.* **2023**, *22*, 895–916. [CrossRef] [PubMed]
42. Kortemme, T. De Novo Protein Design—From New Structures to Programmable Functions. *Cell* **2024**, *187*, 526–544. [CrossRef] [PubMed]
43. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
44. Huerta, E.A.; Khan, A.; Huang, X.; Tian, M.; Levental, M.; Chard, R.; Wei, W.; Heflin, M.; Katz, D.S.; Kindratenko, V.; et al. Accelerated, Scalable and Reproducible AI-Driven Gravitational Wave Detection. *Nat. Astron.* **2021**, *5*, 1062–1068. [CrossRef]
45. López, C. Artificial Intelligence and Advanced Materials. *Adv. Mater.* **2023**, *35*, 2208683. [CrossRef]
46. Wang, H.; Fu, T.; Du, Y.; Gao, W.; Huang, K.; Liu, Z.; Chandak, P.; Liu, S.; Van Katwyk, P.; Deac, A.; et al. Scientific Discovery in the Age of Artificial Intelligence. *Nature* **2023**, *620*, 47–60. [CrossRef]
47. Berry, T.; Harlim, J. Correcting Biased Observation Model Error in Data Assimilation. *Mon. Wea. Rev.* **2017**, *145*, 2833–2853. [CrossRef]
48. Cintra, R.; de Campos Velho, H.; Cocke, S. Tracking the Model: Data Assimilation by Artificial Neural Network. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 403–410.
49. Yuval, J.; O'Gorman, P.A.; Hill, C.N. Use of Neural Networks for Stable, Accurate and Physically Consistent Parameterization of Subgrid Atmospheric Processes with Good Performance at Reduced Precision. *Geophys. Res. Lett.* **2021**, *48*, e2020GL091363. [CrossRef]
50. McGovern, A.; Elmore, K.L.; Gagne, D.J.; Haupt, S.E.; Karstens, C.D.; Lagerquist, R.; Smith, T.; Williams, J.K. Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2073–2090. [CrossRef]
51. Chen, G.; Wang, W.-C. Short-Term Precipitation Prediction for Contiguous United States Using Deep Learning. *Geophys. Res. Lett.* **2022**, *49*, e2022GL097904. [CrossRef]
52. Espeholt, L.; Agrawal, S.; Sønderby, C.; Kumar, M.; Heek, J.; Bromberg, C.; Gazen, C.; Carver, R.; Andrychowicz, M.; Hickey, J.; et al. Deep Learning for Twelve Hour Precipitation Forecasts. *Nat. Commun.* **2022**, *13*, 5145. [CrossRef] [PubMed]
53. Chen, X.; Wang, M.; Wang, S.; Chen, Y.; Wang, R.; Zhao, C.; Hu, X. Weather Radar Nowcasting for Extreme Precipitation Prediction Based on the Temporal and Spatial Generative Adversarial Network. *Atmosphere* **2022**, *13*, 1291. [CrossRef]
54. Chen, Y.; Huang, G.; Wang, Y.; Tao, W.; Tian, Q.; Yang, K.; Zheng, J.; He, H. Improving the Heavy Rainfall Forecasting Using a Weighted Deep Learning Model. *Front. Environ. Sci.* **2023**, *11*, 1116672. [CrossRef]
55. Wang, J.; Wang, X.; Guan, J.; Zhang, L.; Zhang, F.; Chang, T. STPF-Net: Short-Term Precipitation Forecast Based on a Recurrent Neural Network. *Remote Sens.* **2024**, *16*, 52. [CrossRef]
56. Meng, F.; Yang, K.; Yao, Y.; Wang, Z.; Song, T. Tropical Cyclone Intensity Probabilistic Forecasting System Based on Deep Learning. *Int. J. Intell. Syst.* **2023**, *2023*, 3569538. [CrossRef]
57. Wu, Y.; Geng, X.; Liu, Z.; Shi, Z. Tropical Cyclone Forecast Using Multitask Deep Learning Framework. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6503505. [CrossRef]
58. Khodayar, M.; Wang, J.; Manthouri, M. Interval Deep Generative Neural Network for Wind Speed Forecasting. *IEEE Trans. Smart Grid* **2019**, *10*, 3974–3989. [CrossRef]
59. Jiang, S.; Fan, H.; Wang, C. Improvement of Typhoon Intensity Forecasting by Using a Novel Spatio-Temporal Deep Learning Model. *Remote Sens.* **2022**, *14*, 5205. [CrossRef]
60. Chiranjeevi, B.S.; Shreegagana, B.; Bhavana, H.S.; Karanth, I.; Asha Rani, K.P.; Gowrishankar, S. Weather Prediction Analysis Using Classifiers and Regressors in Machine Learning. In Proceedings of the 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 23–25 January 2023; p. 900, ISBN 978-1-66547-467-2.
61. Scher, S. Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model with Deep Learning. *Geophys. Res. Lett.* **2018**, *45*, 12–616. [CrossRef]
62. Dueben, P.D.; Bauer, P. Challenges and Design Choices for Global Weather and Climate Models Based on Machine Learning. *Geosci. Model. Dev.* **2018**, *11*, 3999–4009. [CrossRef]
63. Scher, S.; Messori, G. Weather and Climate Forecasting with Neural Networks: Using General Circulation Models (GCMs) with Different Complexity as a Study Ground. *Geosci. Model. Dev.* **2019**, *12*, 2797–2809. [CrossRef]
64. Weyn, J.A.; Durran, D.R.; Caruana, R. Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height from Historical Weather Data. *J. Adv. Model. Earth Syst.* **2019**, *11*, 2680–2693. [CrossRef]
65. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning Skillful Medium-Range Global Weather Forecasting. *Science* **2023**, *382*, 1416–1421. [CrossRef]
66. He, T.; Yu, S.; Wang, Z.; Li, J.; Chen, Z. From Data Quality to Model Quality: An Exploratory Study on Deep Learning. In Proceedings of the 11th Asia-Pacific Symposium on Internetware, Fukuoka, Japan, 28–29 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–6.

67. Whang, S.E.; Lee, J.-G. Data Collection and Quality Challenges for Deep Learning. *Proc. VLDB Endow.* **2020**, *13*, 3429–3432. [CrossRef]
68. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Quart. J. R. Meteoro Soc.* **2020**, *146*, 1999–2049. [CrossRef]
69. Garg, S.; Rasp, S.; Thuerey, N. WeatherBench Probability: A Benchmark Dataset for Probabilistic Medium-Range Weather Forecasting along with Deep Learning Baseline Models. *arXiv* **2022**, arXiv:2205.00865.
70. Rasp, S.; Hoyer, S.; Merose, A.; Langmore, I.; Battaglia, P.; Russel, T.; Sanchez-Gonzalez, A.; Yang, V.; Carver, R.; Agrawal, S.; et al. WeatherBench 2: A Benchmark for the next Generation of Data-Driven Global Weather Models. *arXiv* **2023**, arXiv:2308.15560.
71. Watson-Parris, D.; Rao, Y.; Olivié, D.; Seland, Ø.; Nowack, P.; Camps-Valls, G.; Stier, P.; Bouabid, S.; Dewey, M.; Fons, E.; et al. ClimateBench v1.0: A Benchmark for Data-Driven Climate Projections. *J. Adv. Model. Earth Syst.* **2022**, *14*, e2021MS002954. [CrossRef]
72. Eyring, V.; Bony, S.; Meehl, G.A.; Senior, C.A.; Stevens, B.; Stouffer, R.J.; Taylor, K.E. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization. *Geosci. Model. Dev.* **2016**, *9*, 1937–1958. [CrossRef]
73. Kurth, T.; Subramanian, S.; Harrington, P.; Pathak, J.; Mardani, M.; Hall, D.; Miele, A.; Kashinath, K.; Anandkumar, A. FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. In Proceedings of the Platform for Advanced Scientific Computing Conference, Davos, Switzerland, 26–28 June 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–11.
74. Chen, K.; Han, T.; Gong, J.; Bai, L.; Ling, F.; Luo, J.-J.; Chen, X.; Ma, L.; Zhang, T.; Su, R.; et al. FengWu: Pushing the Skillful Global Medium-Range Weather Forecast beyond 10 Days Lead. *arXiv* **2023**, arXiv:2304.02948.
75. Chen, L.; Zhong, X.; Zhang, F.; Cheng, Y.; Xu, Y.; Qi, Y.; Li, H. FuXi: A Cascade Machine Learning Forecasting System for 15-Day Global Weather Forecast. *npj Clim. Atmos. Sci.* **2023**, *6*, 190. [CrossRef]
76. Kochkov, D.; Yuval, J.; Langmore, I.; Norgaard, P.; Smith, J.; Mooers, G.; Klöwer, M.; Lottes, J.; Rasp, S.; Düben, P.; et al. Neural general circulation models. *arXiv* **2023**, arXiv:2311.07222.
77. Han, T.; Guo, S.; Ling, F.; Chen, K.; Gong, J.; Luo, J.; Gu, J.; Dai, K.; Ouyang, W.; Bai, L. FengWu-GHR: Learning the Kilometer-Scale Medium-Range Global Weather Forecasting. *arXiv* **2024**, arXiv:2402.00059.
78. Weyn, J.A.; Durran, D.R.; Caruana, R. Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. *J. Adv. Model. Earth Syst.* **2020**, *12*, e2020MS002109. [CrossRef]
79. Rasp, S.; Thuerey, N. Data-Driven Medium-Range Weather Prediction with a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench. *J. Adv. Model. Earth Syst.* **2021**, *13*, e2020MS002405. [CrossRef]
80. Keisler, R. Forecasting Global Weather with Graph Neural Networks. *arXiv* **2022**, arXiv:2202.07575.
81. Nguyen, T.; Brandstetter, J.; Kapoor, A.; Gupta, J.K.; Grover, A. ClimaX: A Foundation Model for Weather and Climate. *arXiv* **2023**, arXiv:2301.10343.
82. Bhardwaj, R.; Duhoon, V. Weather Forecasting Using Soft Computing Techniques. In Proceedings of the 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 28–29 September 2018; pp. 1111–1115.
83. Ling, F.; Ouyang, L.; Larbi, B.R.; Luo, J.-J.; Zhong, X.; Bai, L. Is Artificial Intelligence Providing the Second Revolution for Weather Forecasting? *arXiv* **2024**, arXiv:2401.16669.
84. Olivetti, L.; Messori, G. Advances and Prospects of Deep Learning for Medium-Range Extreme Weather Forecasting. *Geosci. Model. Dev.* **2024**, *17*, 2347–2358. [CrossRef]
85. Cong, S.; Zhou, Y. A Review of Convolutional Neural Network Architectures and Their Optimizations. *Artif. Intell. Rev.* **2023**, *56*, 1905–1969. [CrossRef]
86. Zhou, T.; Ma, Z.; Wang, X.; Wen, Q.; Sun, L.; Yin, T.Y.W.; Jin, R. FiLM: Frequency Improved Legendre Memory Model for Long-Term Time Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 12677–12690.
87. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241, ISBN 978-3-319-24573-7.
88. Schultz, M.G.; Betancourt, C.; Gong, B.; Kleinert, F.; Langguth, M.; Leufen, L.H.; Mozaffari, A.; Stadtler, S. Can Deep Learning Beat Numerical Weather Prediction? *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* **2021**, *379*, 20200097. [CrossRef] [PubMed]
89. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
90. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. [CrossRef] [PubMed]
91. Alet, F.; Jeewajee, A.K.; Villalonga, M.B.; Rodriguez, A.; Lozano-Perez, T.; Kaelbling, L. Graph Element Networks: Adaptive, Structured Computation and Memory. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 212–222.
92. Sanchez-Gonzalez, A.; Godwin, J.; Pfaff, T.; Ying, R.; Leskovec, J.; Battaglia, P. Learning to Simulate Complex Physics with Graph Networks. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 8459–8468.

93. Pfaff, T.; Fortunato, M.; Sanchez-Gonzalez, A.; Battaglia, P.W. Learning Mesh-Based Simulation with Graph Networks. *arXiv* **2020**, arXiv:2010.03409.
94. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
95. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
96. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
97. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling Up Capacity and Resolution. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 12009–12019.
98. Guibas, J.; Mardani, M.; Li, Z.; Tao, A.; Anandkumar, A.; Catanzaro, B. Adaptive Fourier Neural Operators: Efficient Token Mixers for Transformers. *arXiv* **2022**, arXiv:2111.13587.
99. Watt-Meyer, O.; Dresdner, G.; McGibbon, J.; Clark, S.K.; Henn, B.; Duncan, J.; Brenowitz, N.D.; Kashinath, K.; Pritchard, M.S.; Bonev, B.; et al. ACE: A Fast, Skillful Learned Global Atmospheric Model for Climate Prediction. *arXiv* **2023**, arXiv:2310.02074.
100. Bonev, B.; Kurth, T.; Hundt, C.; Pathak, J.; Baust, M.; Kashinath, K.; Anandkumar, A. Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023.
101. Temam, R. *Navier-Stokes Equations: Theory and Numerical Analysis*; American Mathematical Soc.: Washington, DC, USA, 2001; Volume 343, ISBN 0-8218-2737-5.
102. Bauer, P.; Quintino, T.; Wedi, N.; Bonanni, A.; Chrust, M.; Deconinck, W.; Diamantakis, M.; Düben, P.; English, S.; Flemming, J. *The ECMWF Scalability Programme: Progress and Plans*; European Centre for Medium Range Weather Forecasts: Reading, UK, 2020.
103. Lofstead, J. Weather Forecasting Limitations in the Developing World. In *Proceedings of the Distributed, Ambient and Pervasive Interactions*; Streitz, N.A., Konomi, S., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 86–96.
104. Xu, R.; Han, F.; Ta, Q. Deep Learning at Scale on NVIDIA V100 Accelerators. In Proceedings of the 2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS), Dallas, TX, USA, 12 November 2018; pp. 23–32.
105. Jeon, W.; Ko, G.; Lee, J.; Lee, H.; Ha, D.; Ro, W.W. Chapter Six—Deep Learning with GPUs. In *Advances in Computers*; Kim, S., Deka, G.C., Eds.; Hardware Accelerator Systems for Artificial Intelligence and Machine Learning; Elsevier: Amsterdam, The Netherlands, 2021; Volume 122, pp. 167–215.
106. Jouppi, N.; Kurian, G.; Li, S.; Ma, P.; Nagarajan, R.; Nai, L.; Patil, N.; Subramanian, S.; Swing, A.; Towles, B.; et al. TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings. In Proceedings of the 50th Annual International Symposium on Computer Architecture, Orlando, FL, USA, 17–21 June 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–14.
107. Shen, B.-W.; Pielke, R.A.; Zeng, X.; Baik, J.-J.; Faghih-Naini, S.; Cui, J.; Atlas, R. Is Weather Chaotic?: Coexistence of Chaos and Order within a Generalized Lorenz Model. *Bull. Am. Meteorol. Soc.* **2021**, *102*, E148–E158. [CrossRef]
108. McGovern, A.; Lagerquist, R.; Gagne, D.J., II; Jergensen, G.E.; Elmore, K.L.; Homeyer, C.R.; Smith, T. Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 2175–2199. [CrossRef]
109. Liu, Q.; Lou, X.; Yan, Z.; Qi, Y.; Jin, Y.; Yu, S.; Yang, X.; Zhao, D.; Xia, J. Deep-Learning Post-Processing of Short-Term Station Precipitation Based on NWP Forecasts. *Atmos. Res.* **2023**, *295*, 107032. [CrossRef]
110. Hewamalage, H.; Ackermann, K.; Bergmeir, C. Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices. *Data Min. Knowl. Discov.* **2023**, *37*, 788–832. [CrossRef] [PubMed]
111. Bouallègue, Z.B.; Clare, M.C.A.; Magnusson, L.; Gascón, E.; Maier-Gerber, M.; Janoušek, M.; Rodwell, M.; Pinault, F.; Dramsch, J.S.; Lang, S.T.K.; et al. The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning-Based Weather Forecasts in an Operational-like Context. *Bull. Am. Meteorol. Soc.* 2024; *published online ahead of print*. [CrossRef]
112. Saleem, H.; Salim, F.; Purcell, C. Conformer: Embedding Continuous Attention in Vision Transformer for Weather Forecasting. *arXiv* **2024**, arXiv:2402.17966.
113. Ding, N.; Qin, Y.; Yang, G.; Wei, F.; Yang, Z.; Su, Y.; Hu, S.; Chen, Y.; Chan, C.-M.; Chen, W.; et al. Parameter-Efficient Fine-Tuning of Large-Scale Pre-Trained Language Models. *Nat. Mach. Intell.* **2023**, *5*, 220–235. [CrossRef]
114. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models Are Few-Shot Learners. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 1877–1901.
115. Shawki, N.; Nunez, R.R.; Obeid, I.; Picone, J. On Automating Hyperparameter Optimization for Deep Learning Applications. In Proceedings of the 2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 4 December 2021; pp. 1–7.

116. Park, H.; Nam, Y.; Kim, J.-H.; Choo, J. HyperTendril: Visual Analytics for User-Driven Hyperparameter Optimization of Deep Neural Networks. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 1407–1416. [CrossRef]
117. Bergstra, J.; Yamins, D.; Cox, D. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. In Proceedings of the 12th Python in Science Conference, Austin, TX, USA, 24–29 June 2013; pp. 13–19.
118. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 2623–2631.
119. Golovin, D.; Solnik, B.; Moitra, S.; Kochanski, G.; Karro, J.; Sculley, D. Google Vizier: A Service for Black-Box Optimization. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1487–1495.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.